

Bringing archival collections to Wikipedia with the Remixing Archival Metadata Project (RAMP) editor

Mairelys Lemus-Rojas, Metadata Librarian
Timothy A. Thompson, Discovery Metadata Librarian

Wikipedia and archives

Wikipedia has been in existence for 16 years, but it was not until recently that GLAM (Galleries, Libraries, Archives, and Museums) institutions started to see it as a trusted source and one they were interested in contributing to. Libraries and archives, in particular, have been slowly embracing the idea of contributing to Wikipedia, often motivated by the prospect of increasing the visibility and impact of their distinctive collections, which have been carefully curated by librarians and archivists alike. Different approaches have been taken by libraries and archives when it comes to contributing to Wikipedia. Some institutions have focused on adding links to Wikipedia articles in order to point back to their unique archival collections, whereas others have concentrated on enhancing the content of existing articles. Another way in which librarians and archivists can work together to share our collections is to collaborate on projects to enrich and repurpose the metadata that has already been created to describe them.

The University of Miami Libraries took this approach and developed a tool to facilitate the creation of Wikipedia articles using relevant metadata from the libraries' finding aids (documents that describe the scope, content, and context of archival collections). In this chapter, we will introduce the RAMP (Remixing Archival Metadata Project) editor and share our experience working on a pilot project conducted to test its viability.

Background

The RAMP editor (<https://tools.wmflabs.org/ramp/>) was developed to provide the cultural heritage community with the ability to reuse, remix, and republish curated metadata via the English Wikipedia. RAMP is a web-based editing tool that allows users to generate records for creators of archival collections and publish the content as Wikipedia pages. It is built around two metadata formats, Encoded Archival Description (EAD) and Encoded Archival Context—Corporate Bodies, Persons, and Families (EAC-CPF).

These formats have been developed to provide organized access to archival and special collections. Findings aids, typically encoded using the EAD XML schema, now in its third edition (Library of Congress, 2017), are hierarchically structured documents that reflect the arrangement and disposition of a collection within an archive. EAD-encoded finding aids can be complex or simple, depending on the time and resources available for metadata creation, but they usually contain a section of contextual information (the scope of the collection, the biography or history of its creator), and a list of the collection's contents, often divided into series that reflect the provenance and original or intellectual order of the material.

Compared to EAD, EAC-CPF is a newer standard, and it has been codified in a separate XML schema (Staatsbibliothek zu Berlin, n.d.). The importance of context in understanding and maintaining archival collections entails an emphasis on relationships, whether among collections or the individuals, families, and organizations responsible for creating their content. EAC-CPF was designed to serve as a kind of authority record for creators of archival collections, but its purpose extends beyond recording variant names, for example, and provides a vehicle for

recording the many collections that might contain works by a single creator (for instance, a literary author's papers might be held in multiple, geographically dispersed repositories), as well as relationships to other creators and content.

Building the RAMP tool was just a small step toward providing the cultural heritage community with a way to more easily repurpose and expose its archival metadata through a widely accessed platform, Wikipedia. For some, the notion of contributing to Wikipedia is challenging and even discouraging, and the idea that the global community rather than a formalized entity is curating and contributing to this body of knowledge is a concept that has not yet been fully embraced by information professionals. Wikipedia allows for many voices, and we should actively join the conversation rather than passively observing its evolution.

Original tool development

The original development of the tool started in the early 2013 and took approximately two months. This was an in-house project, and no external funding was used toward its development.

RAMP is an open source tool, and its code is available on GitHub

(<https://github.com/UMiamiLibraries/RAMP>) for the community to use and modify. This project presented an opportunity for cross-departmental collaboration at the University of Miami Libraries and brought together librarians, archivists, and programmers working toward a common goal.

The original version of RAMP was developed using the standard LAMP stack of Linux, Apache, MySQL, and PHP, with a JavaScript front end. Its architecture was based on a software design pattern known as Model-View-Controller (MVC). On the back end, the PHP programming

language is used to model the basic structure of the tool and to interact directly with three main web services: the Virtual International Authority File (VIAF), WorldCat Identities, and Wikipedia.¹ Data from these services is processed using JavaScript, then returned to the user.

At the heart of the design are two data transformation routines executed by a suite of XSLT (Extensible Stylesheet Language Transformations) stylesheets. Raw data is ingested as EAD/XML, and biographical information and information about significant relationships is extracted from the EAD and converted into EAC-CPF/XML, which is then exposed for enhancement through data lookups or direct user editing. A second XSLT transformation converts the enhanced EAC-CPF from XML to wikitext for publication to Wikipedia. The wikitext can be edited, saved, and merged with existing Wikipedia pages within the tool.

A raw XML editing interface for EAC-CPF records is provided through integration with the code editing software Ace Editor (Ajax.org, 2017). Although not ideal from a user-interface perspective, this basic editing capability proved to be a lightweight solution for viewing the results of the data enhancement process. In the original RAMP design, a web form was also included to allow users to create new EAC-CPF records. Later, a decision was made not to

¹ VIAF (<http://viaf.org/>) and WorldCat Identities (<http://worldcat.org/identities/>) are both OCLC-hosted open data aggregating services. VIAF merges data from significant international name authority files and mints a single URI for unique identity clusters. WorldCat Identities aggregates data from across the WorldCat database to produce a single view of information related to personal, corporate, and subject-based identities, including works by, works about, associated names, and related authorities (OCLC Research, 2012).

continue supporting this feature because a robust EAC-CPF creation tool, the xEAC editor, was already available and under active development.²

RAMP pilot project

To explore the potential of the RAMP tool, we conducted a pilot project in the spring of 2014.

We were interested in finding out how long it would take to create a new Wikipedia article and whether this process was something that could be easily integrated into existing workflows without impinging on other responsibilities. We were also curious to find out whether our contributions to the encyclopedia would increase web traffic to the library website.

Once we had established what we were interested in testing, we initiated a conversation with stakeholders from the University of Miami's Cuban Heritage Collection, who selected a set of collections to be used in the pilot project. These collections represented individuals and organizations active in Cuban theater circles, and the focus of their content was not well represented in the English Wikipedia. We started by reviewing a set of finding aids to identify potential candidates for inclusion in the encyclopedia. Our first criterion was to select those that had the most complete biographical notes.

² xEAC (pronounced "zeek") is an open source EAC-CPF editor created by Ethan Gruber using the declarative programming language XForms. It provides both front-end and administrative interfaces, as well as integration with a range of web service APIs, including SNAC, DBPedia, VIAF, Getty AAT, Getty TGN, Geonames, the Pleiades Gazetteer of Ancient Places, and the LC Name Authority File for geographic places (Gruber, 2017).

Next, we considered the “notability” of each subject. Wikipedia guidelines require that subjects be notable in order to merit inclusion in the encyclopedia. The general notability guideline states that “if a topic has received significant coverage in reliable sources that are independent of the subject, it is presumed to be suitable for a stand-alone article or list” (“Wikipedia:Notability,” 2017). Decisions about whether a subject is notable or not are often controversial and have been a sticking point in the past for librarians and archivists interested in contributing to Wikipedia (Tennant, 2013). To avoid potential conflict, we selected what we considered to be the most “notable” collections for the pilot project: a total of 18 out of the 32 available Cuban theater collections. These collections represented Cuban and Cuban American actors, playwrights, directors, costume designers, and theater companies.

Although we identified collections that had already been processed, we intended to continue the work beyond the pilot project. We devised a workflow that would function with new collections and would involve key players from various departments. Once a collection was processed and a finding aid was created, the archivist working with the collection would submit a work request via a web form. Librarians in Cataloging & Metadata Services would receive an email with the details of the work request. At that point, a librarian would claim the task and start creating a collection-level record and related authority records in OCLC Connexion, which would later be pushed to our local integrated library system. Finally, the same librarian would begin working with the RAMP tool to create Wikipedia articles for selected subjects.

For the pilot project, because the finding aids had already been created, we followed a modified workflow. The finding aids had been written by many different staff members over the years,

and we wanted to be certain they all reflected a neutral point of view.³ In republishing biographical descriptions from finding aids, we tried to avoid “peacock terms” (“Wikipedia:Manual of Style/Words to watch,” 2017) and removed language that described subjects in non-neutral terms such as a “versatile artist” or someone who had “devoted her life” to a pursuit. We received assistance from a staff writer in the library’s Communications Department, who reviewed the finding aids and suggested edits. After she had finished revising them, we received an email with the updated version of the text. We divided the finding aids between ourselves, but did not start working with them until we had become familiar with the culture and core principles of the Wikipedia editing community.

Once we started using the RAMP tool, we asked ourselves several questions when the results did not match what we expected to find. For instance, when we retrieved data from VIAF and received no results, this meant that no authority record had been created for a particular person or organization. The absence of an authority record reinforced the importance of providing a presence in Wikipedia for these underrepresented creators. It also allowed us to reflect on the value of creating and contributing authority records so that tools like RAMP could be used to share information in meaningful ways. Once we had completed the steps in the editing process and converted our EAC-CPF records into wikitext, we were ready to publish to Wikipedia. The

³ The concept of neutrality is another foundational principle of Wikipedia. Wikipedia’s neutral point of view guidelines state that “all encyclopedic content on Wikipedia must be written from a neutral point of view (NPOV), which means representing fairly, proportionately, and, as far as possible, without editorial bias, all of the significant views that have been published by reliable sources on a topic” (“Wikipedia:Neutral point of view,” 2017).

tool provides the option to contribute information directly to a live Wikipedia page (in cases where data is used to enhance an existing article) or as a draft of a new article, created as a subpage under the Wikipedia editor's user page. Because 17 of the articles were new to Wikipedia, we relied primarily on the draft article workflow, but tested making live edits as well.

The backbone of a Wikipedia article is built on citations and links to external resources, which provide users with the necessary information to continue their research and validate the reliability of an article. We included links under the "External links" section (see Figure 1) of each article, which referred readers back to the finding aid and any related digital collections. We also included a link to the library website and whenever available, we provided a link to the subject's respective page in the Cuban Theater Digital Archive (CTDA) website, a thematic database curated by University of Miami faculty member Lillian Manzor (Cuban Theater Digital Archive, 2011). We were careful to link to resources that we believed would be reliable and relevant to researchers.

External links [edit]

- The [Caridad Svich papers](#) are available at the [Cuban Heritage Collection](#), [University of Miami Libraries](#). This archival collection contains the personal and literary works of playwright Caridad Svich. The collection is organized into two series. Series 1 consists of original theater works as well as translations by Svich; materials related to theater productions, such as playbills, advertisements, clippings, and reviews; and articles about other playwrights. Series 2 consists of audiovisual material such as MiniDV's, CDs, and tapes containing presentations by Caridad Svich and performance soundtracks.
- Selected photographs from the [Caridad Svich papers](#) are available through the [University of Miami Libraries Digital Collections](#) portal.
- Creator page for [Caridad Svich](#) in the [Cuban Theater Digital Archive](#).

Authority control [WorldCat Identities](#) • [VIAF](#): [74098334](#) • [LCCN](#): [no98080414](#) • [SUDOC](#): [093926928](#) • [BNF](#): [cb14582433s](#) [\(data\)](#)

Figure 1. External links section of a Wikipedia article created using RAMP.

RAMP assists editors in improving some of the key features of a Wikipedia article. For example, the tool can help supply basic data for the article's infobox⁴, such as birth and death dates when available (see Figure 2). The biographical/historical section of the Wikipedia article is drawn from the biographical/historical note in the collection's finding aid. RAMP-created articles also include a "Works or publications" section that contains data imported from WorldCat Identities during one of the processes run within the tool. RAMP adds a hidden category to each Wikipedia article that allows articles created using the tool to be tracked ("Category:Articles with information extracted by the RAMP editor," 2017).

Caridad Svich	
Born	July 30, 1963 (age 53) Philadelphia, Pennsylvania
Nationality	American
Education	University of North Carolina at Charlotte (BFA) University of California, San Diego (MFA)
Occupation	Playwright
Awards	2012 Obie Award for Lifetime Achievement

Figure 2. Infobox of an article created using RAMP.

⁴ An infobox contains structured metadata pertaining to the subject being described in the Wikipedia article. It is usually displayed on the right-hand side of an article for languages that read from left to right.

We spent roughly 25 hours working on the pilot project. This total included time spent assessing the 32 theater collections and selecting which to include in Wikipedia, based on completeness and notability.

Rights Statement

During the planning stages of the pilot, we looked at Wikipedia's requirements for contributing data. Because we were copying text directly from the finding aids, it was necessary to indicate that the text we were repurposing had been appropriately licensed and could be copied verbatim. We added a rights statement to each of the finding aids we had selected, following Wikipedia guidelines ("Wikipedia:Reusing Wikipedia content," 2017). We added a Creative Commons Attribution-ShareAlike 3.0 license and GNU Free Documentation license to the finding aid pages, placing them in an "Other Note" field because there was no more specific field in the EAD schema to store this information. In the Wikipedia pages, we included a similar rights statement under the "Notes and references" section, with our citation of the finding aid document.

Outcomes

Bringing diversity to Wikipedia

The need for greater diversity on Wikipedia, both in terms of content and the composition of the editing community, has been the target of much critical commentary. This focus has resulted in thematic edit-a-thons such as those organized by participants in the Art+Feminism collective (Art+Feminism, n.d.). Similarly, library and archival collections and the profession as a whole face their own issues with diversity. Important work is being done to critique and question the

role of archives in perpetuating exclusionary norms that continue to privilege heterosexual white men. As Jarrett M. Drake (2016) states:

I'm skeptical about archives in the United States. Even more specifically, I'm skeptical about archives in the United States that adhere to the standard tradition of archives in the Western world. I've spent hours and hours of my time this year and last reading, thinking, tweeting, and writing about the origins of my skepticism, while also reconciling what it means that I am so very much a part of the problem that I see in this work and trying to advocate for the abolition of the archaic, anti-black, transphobic, elitist and misogynistic aspects of archival administration.

Tools such as RAMP can be used to draw attention to both sides of this deficit of diversity and inclusion. Collections that showcase the work and accomplishments of women or minorities can and should be prioritized for “wikifying” in order to help fill important gaps and ensure that standards of notability do not simply reinforce existing patterns of oppression and inequality.

Measuring web traffic to finding aid pages

In 2011 a project was carried out at Ball State University to update relevant Wikipedia articles with links to digital assets in the university's Hague Sheet Music Collection (Szajewski, 2013). Web traffic captured by Google Analytics was analyzed for the year preceding and following the addition of links, and it was found that in the subsequent year, pageviews of digital assets in the

collection had increased by 610.31% (from 1,824 to 12,956), with Wikipedia accounting for nearly 76% of web traffic referrals.⁵

We undertook a similar analysis to measure the impact of adding backlinks from the Wikipedia pages we had created to their corresponding University of Miami finding aid pages. We produced a Google Analytics report for traffic to the finding aid website for the two-year period between April 2013 and April 2015. Incoming links from the University of Miami network domain were excluded from the report in order to minimize bias and eliminate traffic that had been incidentally generated by participants in the pilot project. This may have excluded legitimate traffic from other on-campus users, but we chose to limit our focus to external users for the purposes of this analysis. Two separate datasets were created: one for the finding aids used in the pilot project and one for all other finding aid pages, used as a control group. Project pages were identified by matching the collection identifier (included in the finding aid URL) against the Google Analytics field that contained the relative URL path of the landing page for each browsing session.

Subsequently, the creation date of each Wikipedia page was added to the entries in the pilot project Analytics dataset. This date was used to create a before-and-after marker so that the impact of linking from Wikipedia could be evaluated. All but one of the finding aid pages had

⁵ According to Google's documentation, a pageview "is defined as a view of a page on your site that is being tracked by the Analytics tracking code. If a user clicks reload after reaching the page, this is counted as an additional pageview. If a user navigates to a different page and then returns to the original page, a second pageview is recorded as well" (Google, 2017a).

received some traffic in the post-Wikipedia period, and total web browsing sessions had increased by 39% (from 505 to 701). A single browsing session can represent multiple pageviews, and there was an overall increase in pageviews of 36% (from 682 to 930). Before the pilot project, only one of the finding aids had been linked to from a Wikipedia page (with a total of 2 pageviews). After the pilot project, links from Wikipedia accounted for 21% of all pageviews for the pilot project finding aid pages (192 out of 930).

The control dataset, which recorded web traffic to all finding aid pages except the ones included in the pilot project, also showed an increase over time, albeit a slightly smaller one. Total browsing sessions increased by 32% in the second period (from 10,861 to 14,331), and pageviews increased by 23% (from 28,856 to 35,449). In the period preceding the pilot project, Wikipedia accounted for about 2% of all pageviews in the control dataset (626 out of 28,856 total pageviews), and in the subsequent period, it again accounted for approximately 2% of pageviews (658 out of 35,449).

The pilot project data was inspected in greater detail using a one-way analysis of variance (ANOVA) procedure.⁶ ANOVA results indicated that the difference between the mean value of pageviews for the two periods was not statistically significant at the 0.05 level ($p = 0.2347$). Although the frequency of pageviews had increased over time, the average number of pageviews

⁶ Statistical analyses were performed using the SAS 9.4 software package. Analysis of variance was applied using the GLM procedure. Levene's test for homogeneity of variance was not significant at the 0.05 level ($p = 0.2347$), indicating that analysis of variance was an appropriate statistical method for the data.

had remained approximately the same: the mean number of pageviews in the pre-Wikipedia period was 1.35, versus 1.33 in the subsequent period (with a minimum significant difference of 0.107). Grouping the pageviews by month yielded some separation between the mean values for the two periods: the pre-Wikipedia mean number of pageviews per month was 38.85, and the post-Wikipedia mean was 50.07, but the difference was still not statistically significant at the 0.05 level ($p = 0.2869$, with a minimum significant difference of 21.246).

The impact of linking from Wikipedia was also examined using a difference-in-differences analysis of the control and pilot datasets.⁷ To make the number of observations equal for both datasets, a random sample was taken from the control data for each time period (see Table 1 and Figure 3). The following regression model was tested:

$$\text{Pageviews} = \beta_1 + \beta_2\text{Group} + \beta_3\text{Period} + \beta_4(\text{Group})(\text{Period}) + u_i$$

where Group represents the categorical variable for the control and pilot groups and Period represents the categorical variable for the pre- and post-Wikipedia periods. A statistically significant value for the parameter β_4 would suggest that the addition of Wikipedia backlinks had led to an increase in web traffic to the finding aid pages.

⁷ Difference-in-differences is a quasi-experimental statistical method used to determine whether a particular treatment, intervention, or policy has had an impact over time. Binary variables are used to divide the data into two cross-sectional groups, treatment and control, and to differentiate between the before and after periods for each. Analysis of the interaction term (treatment and time period) indicates whether the intervention has had a statistically significant effect.

	Control group		Pilot group	
	Before (n = 505)	After (n = 701)	Before (n = 505)	After (n = 701)
Count	1445	2088	682	930
Mean	2.86	2.98	1.35	1.33
Std	4.63	6.56	0.77	1.04

Table 1. Summary statistics for pageviews by cross-section and time period.

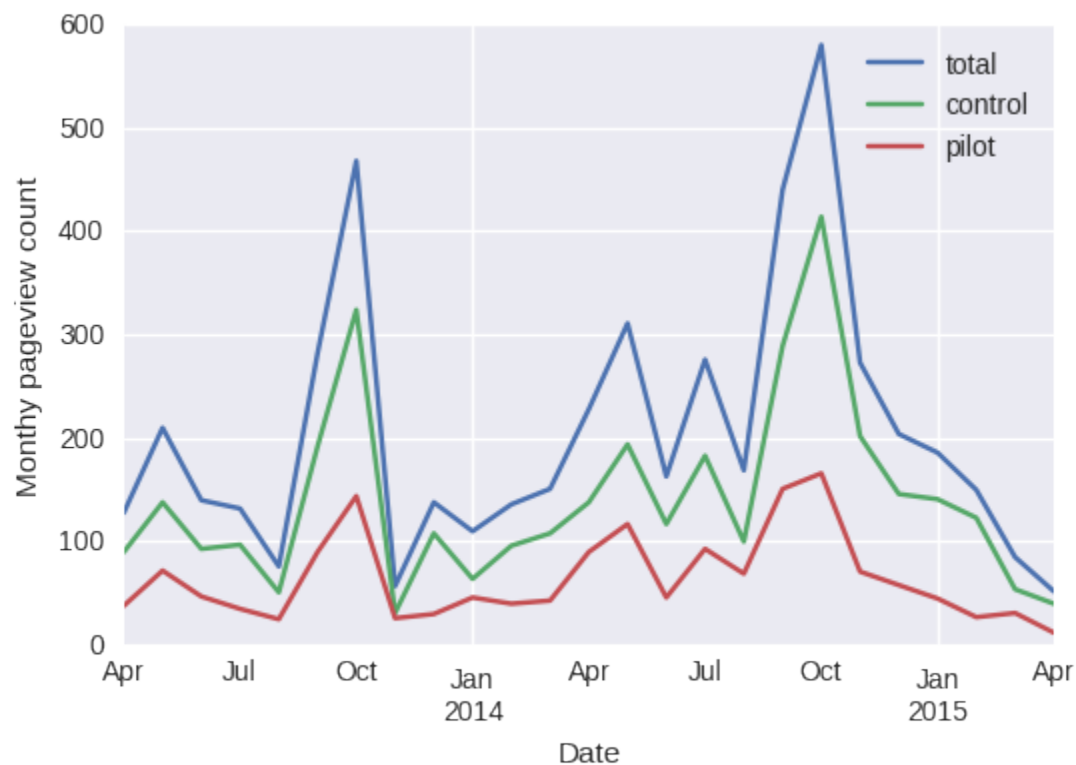


Figure 3. Monthly pageviews by cross-section.

Although the overall model was statistically significant ($F = 29.33$, $p = < 0.0001$), its practical significance was minimal. The adjusted R^2 value for the model was 0.0341, suggesting that only 3% of the variation in the number of pageviews could be attributed to changes in the independent variables, after adjusting for the number of independent variables. Values for the β_4 parameter indicated that there was no statistically significant evidence of a relationship between web traffic and the addition of backlinks ($t = -0.41$, $p = 0.6824$), holding the effects of the other independent variables constant. The only significant independent variable was the Group variable, which distinguished between control and pilot pages over both time periods. Results indicated that membership in the pilot group corresponded to a decrease in pageviews of 1.51089, holding the effects of the other variables constant.

Based on the Google Analytics data presented here, it can be concluded that the pages selected for the pilot project were already less likely to be viewed than other University of Miami finding aid pages. The numerically low average for pilot group pageviews over both periods suggests that the pilot pages were more likely to be of interest to specialized researchers than the general public. The choice of subject matter for the pilot project (Cuban theater collections) was a factor that defined the scope of the project's prospective appeal, limiting it to a particular audience: namely, people who followed or studied theater generally and Hispanic American or Latin American theater specifically. In comparison, the Ball State project cited above involved adding backlinks to Wikipedia pages about well-known nineteenth- and twentieth-century popular songs such as "It's a long, long way to Tipperary" (Szajewski, 2013). The subject matter selected in that case was one that already had the potential to appeal to a large audience of prospective users.

A comparison of the two projects (the University of Miami RAMP pilot and the Ball State project) highlights the tension between notability and visibility. The concept of notability seems beset by a catch-22: in order to be notable, a subject must first be visible, but in order to become visible, a subject must first be notable. Projects such as RAMP might be seen as a first step toward asserting notability through visibility. The process of actively constructing notability (rather than allowing it to limit the scope of our contributions) is part of a long-term effort to enrich the broader information ecosystem and, whenever possible, draw attention to traditionally marginalized and underrepresented content and content creators. The temptation to emphasize short-term goals such as boosting web traffic is one that should be resisted.


Google Knowledge Graph


More difficult to quantify, but easier to observe is the impact that having a Wikipedia page had on the web presence of the individuals and organizations that were the focus of the pilot project. Prior to the pilot project, information about creators of the collections we had selected was dispersed on the web. The results of the project highlighted the fact that the creation of a Wikipedia page is really the creation of a knowledge hub that helps organize and aggregate related online information. This effect is seen most notably in the example of Google's Knowledge Graph cards (Google, 2017b). For instance, a Google search for "Caridad Svich," a playwright who lacked a Wikipedia page before the project, now displays a prominent knowledge card: key biographical facts have been extracted from her Wikipedia article, combined with photographs and relationships ("People also search for") aggregated from the wider web (see Figure 4). In effect, the creation of a Wikipedia page entails the concomitant

creation of a Google Knowledge Graph card.

Caridad Svich

Playwright



 caridadsvich.com

Caridad Svich is an award-winning playwright, songwriter/lyricist, translator, and editor who was born in the United States of Cuban-Argentine-Spanish-Croatian parents. [Wikipedia](#)

Born: July 30, 1963 (age 53), Philadelphia, PA


Plays: [Luna Park](#)

Education: University of California, San Diego (MFA), University of North Carolina at Charlotte (BFA)


Awards: [Obie Award for Lifetime Achievement](#)

Books


[View 25+ more](#)




Antigone Project: A Play in 5...
2009




Alchemy of Desire/De... Blues
2004



Twelve Ophelias (a Play with...
2008




Theatre in Crisis?
2002




12 Ophelias (a Play with...
2008


People also search for



Tanya Barfield



Julia Alvarez



Jane Shaw

Figure 4. Google Knowledge Graph card display for Caridad Svich (captured June 1, 2017).

Lessons Learned

Despite our efforts to familiarize ourselves with and abide by Wikipedia community norms, there were still some lessons to be learned. In one instance, the notability of one of our subjects was questioned by another Wikipedia editor (“Talk:Teatro Avante,” 2014). The page that we had created was for Teatro Avante, a theater organization located in the Little Havana neighborhood of Miami. The editor argued that the article should be deleted because “the subject matter, a theater group, lacks notability. The article actually seems to be spam to support a software product called RAMP.” We promptly responded, making the case that Teatro Avante was “a notable part of the cultural landscape of Miami’s Little Havana neighborhood” and had been the recipient of funding from the National Endowment for the Arts. In addition, we clarified that RAMP was simply a tool designed to help librarians and archivists contribute to Wikipedia.⁸ The issue was quickly resolved when a Wikipedia administrator declined the request to delete the article, accepting our argument for the subject’s notability and agreeing that the content of the article was not promotional.

⁸ On Wikipedia, discussions about the content of an article typically take place on its corresponding “Talk” page, which provides a public forum in which different views can be expressed. This exchange did lead us to make changes to one of the templates we had been using in the project. Previously, we had included a small RAMP icon at the start of our finding aid citation under “Notes and references.” We removed this icon to avoid the appearance of branding.

This minor dispute gave us a fresh perspective of the range of interactions we could expect to have with other editors. Cataloging and metadata librarians do not typically need to think about questions of notability in their daily work, nor do they interact with the community when creating descriptive metadata for library or archival resources. We could see here where the two communities (libraries and Wikipedia) diverged somewhat: one was more focused on following specific rules and using controlled vocabularies to describe and deliver information resources, whereas the other was more focused on determining whether a subject was notable enough to deserve a place in the expanding encyclopedia.

Some of our articles were tagged as orphans, which meant that no links had been added from already existing Wikipedia pages to these new articles. Often, the Wikipedia editor who tagged the articles as orphans also provided suggestions about the steps we could take to address the issue. We then worked to provide access from other Wikipedia pages back to our newly created ones. One perhaps obvious takeaway from the project was the realization that finding aid biographies have not been written with Wikipedia in mind. The project gave us the opportunity to reflect on ways to move beyond traditional archival standards and think about positioning our descriptive metadata in a way that could appeal to a wider audience of information consumers.

Usability testing

In order for other institutions to use RAMP, we knew that the installation process needed to be simplified, especially for small institutions with minimal IT support. In 2016 a round of usability testing was conducted to identify whether there were obstacles that would prevent the community from using and implementing the tool. Seven people (including library staff

members, librarians, and archivists from various departments) agreed to participate. A Google form was created to track responses and suggestions from participants. The testing revealed that the user interface needed to be improved and that users needed an easy way to ingest and delete files. The need for an easier installation process was also confirmed. This information helped generate user stories that were employed as the basis for a subsequent two-week development sprint.

Second iteration of the RAMP tool

The primary issue addressed in the second round of development was the ability to install a local instance of RAMP without any assistance. Installation is now a three-click process and takes only a few minutes to complete. Another major improvement in the tool centered on the user interface. The previous version was text heavy, and users would get lost when reading the instructions. This was addressed with a complete overhaul of the interface, which gave the tool a new look and feel. The process of logging into Wikipedia was also improved. Programmers ensured that all connections to Wikipedia were secure using the https protocol. We also explored using the OAuth standard to manage Wikipedia access (“OAuth,” 2017), but decided not to pursue it because its implementation would have been too complex given the limited timeframe of the sprint and the resources that were available for development. In the first iteration of RAMP, files needed to be ingested by creating a folder on the server side. Now users can ingest a single file directly into the database. The ability to delete files was also added, and JavaScript features were rewritten to conform to more modern single-page web application patterns, using a combination of JQuery and Underscore.js. Finally, maintenance event elements were added to the EAC/XML to record each action performed on a file.

Conclusions

In the library and archives communities we describe resources using standards and controlled vocabularies that in theory should facilitate the discovery and retrieval of information. Using these vocabularies ensures that resources are described in a consistent way, but it may also limit their findability because users are not always familiar with our terminology. We might be more successful in reaching a wider audience if we began to map our vocabularies and standards to the language of the wider web and disseminated even a portion of our information and content through a platform such as Wikipedia.

Our experience using the RAMP tool has convinced us that contributing to Wikipedia and engaging with the Wikipedia community is an important and even necessary activity, one that is directly relevant to our work as metadata librarians. The process of transforming and repurposing the rich contextual information often contained in finding aids forced us to think beyond traditional boundaries and reimagine the ways in which metadata about archival collections could be shared and presented. It also compelled us to grapple with fundamental questions about notability and visibility, questions that matter deeply to archivists as well. How do we decide what is worth archiving, describing, and preserving? How does the metadata we create reflect our professional values? As information professionals, the more we begin to do our work in public, on platforms such as Wikipedia, the more we will be challenged to see ourselves and our institutions as part of a shared, if contested, space of knowledge curation and circulation. We cannot expect to remain relevant as a profession unless we begin to think about how our metadata standards can be made to speak the language of the web, which is increasingly the language of our users.

References

- Ajax.org. (2017, May 27). Ace [Github repository]. Retrieved June 2, 2017, from <https://github.com/ajaxorg/ace>
- Art+Feminism. (n.d.). Retrieved June 2, 2017, from <http://www.artandfeminism.org/>
- Category:Articles with information extracted by the RAMP editor. (2017). In *Wikipedia*. Retrieved June 2, 2017, from https://en.wikipedia.org/wiki/Category:Articles_with_information_extracted_by_the_RAMP_editor
- Cuban Theater Digital Archive. (2011). Retrieved June 2, 2017, from <http://www.cubantheater.org/>
- Drake, J. M. (2016, October 22). Liberatory archives: towards belonging and believing (part 1). In *On Archivy*. Retrieved June 2, 2017, from <https://medium.com/on-archivy/liberatory-archives-towards-belonging-and-believing-part-1-d26aeb0edd1>
- Google. (2017a). The difference between AdWords Clicks, and Sessions, Users, Entrances, Pageviews, and Unique Pageviews in Analytics. In *Analytics Help: Acquisition Reports*. Retrieved June 2, 2017, from <https://support.google.com/analytics/answer/1257084>
- Google. (2017b). Suggest a change to a Knowledge Graph card. In *Google Search Help*. Retrieved June 2, 2017, from <https://support.google.com/websearch/answer/6325583>
- Gruber, E. (2017, January 13). xEAC [Github repository]. Retrieved June 2, 2017, from <https://github.com/ewgl18/xEAC>
- Library of Congress. (2017, March 9). *Encode Archival Description: official site*. Retrieved June 2, 2017, from <https://www.loc.gov/ead/>

OAuth. (2017). In *Wikipedia*. Retrieved June 2, 2017, from <https://en.wikipedia.org/wiki/OAuth>

OCLC Research. (2012, July 23). WorldCat Identities. Retrieved June 2, 2017, from

<http://www.oclc.org/research/themes/data-science/identities.html>

Staatsbibliothek zu Berlin. (n.d.). Encoded Archival Context: Corporate Bodies, Persons, and

Families. Retrieved June 2, 2017, from <http://eac.staatsbibliothek-berlin.de/about.html>

Szajewski, M. (2013). Using Wikipedia to enhance the visibility of digitized archival assets. D-

Lib Magazine, 19(3/4). doi:10.1045/march2013-szajewski

Talk:Teatro Avante. (2017). In *Wikipedia*. Retrieved June 2, 2017, from

https://en.wikipedia.org/wiki/Talk:Teatro_Avante

Tennant, R. (2013, October 24). The winter of Wikipedia's discontent [Blog post]. In *The Digital*

Shift. Retrieved June 02, 2017, from [http://www.thedigitalshift.com/2013/10/roy-tennant-](http://www.thedigitalshift.com/2013/10/roy-tennant-digital-libraries/winter-wikipedias-discontent/)

[digital-libraries/winter-wikipedias-discontent/](http://www.thedigitalshift.com/2013/10/roy-tennant-digital-libraries/winter-wikipedias-discontent/)

Wikipedia:Manual of Style/Words to watch. (2017). In *Wikipedia*. Retrieved June 2, 2017, from

https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Words_to_watch

Wikipedia:Neutral point of view. (2017). In *Wikipedia*. Retrieved June 2, 2017, from

https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view

Wikipedia:Notability. (2017). In *Wikipedia*. Retrieved June 2, 2017, from

<https://en.wikipedia.org/wiki/Wikipedia:Notability>

Wikipedia:Reusing Wikipedia content. (2017). Retrieved June 2, 2017, from

https://en.wikipedia.org/wiki/Wikipedia:Reusing_Wikipedia_content